

The Ghost in AI: Human Error, the Illusion of Machine Agency, and the Specter of Extinction Risk

By Michael M. Nikolettseas

The proposition that artificial intelligence (AI) poses an existential risk—potentially exterminating humanity due to perceived inferiority in computational tasks—has ignited global concern, amplified by warnings from figures like Stephen Hawking, Elon Musk, and industry leaders (Roose, 2023; Bostrom, 2002). In **The Ghost in AI** (2025), I argue that this fear misidentifies the true threat: human civilization’s reliance on a “fundamental error”—the flawed causal narrative of a non-corporeal “self,” rooted in Aristotelian animism and ignorance of spontaneous neural activity. This error, which portrays humans as ghost-like beings inhabiting flesh, parallels the modern misattribution of agency to AI, creating a “ghost in AI.” Drawing on neuroscience, my statistical approach to causality from **Statistics: Concepts and Examples** (2014), and philosophical insights from Heraclitus and Parmenides, I contend that AI lacks true agency, and the extinction risk narrative reflects human anthropomorphism, not machine intentionality. This article expands on these ideas, integrating the neuroscience of spontaneous neural activity, the concepts of “I,” “self,” and agency, and a critical examination of AI extinction risk, arguing that the true danger lies in human civilization’s persistent causal misattributions.

The Ghost Civilization: A Legacy of Causal Error

In **Behavioral and Neural Plasticity** (2010), I proposed that the human “I” and “self” are non-corporeal fictions, invented to attribute behaviors to a causal agent when the brain’s spontaneous neural activity—neurons firing without external stimuli—is unknown. **The Ghost in AI** extends this, asserting that Earth hosts “over 8 billion ghosts who possess flesh and bones,” living miserable lives and destroying nature due to an Aristotelian concept of self, rooted in animism (Nikolettseas, 2025, p. 1). This “ghost civilization” stems from a historical failure to overcome the limitations of natural language and causal attribution.

Ancient philosophers Heraclitus and Parmenides recognized that natural language inadequately describes reality, necessitating a new method or “calculus” to uncover truth (Nikolettseas, 2015). Plato and Aristotle, however, remained trapped in linguistic frameworks, perpetuating animistic notions of agency. Aristotle’s “prime mover”—the idea of an agent causing events—dominated for two millennia until Galileo introduced empirical science (Nikolettseas, 2025, p. 1). Despite scientific progress, our concept of self remains Aristotelian, viewing the “I” as a non-physical agent commanding the body.

For example, when you grasp a coffee cup and claim “I” caused the action, you imply your “I” exists outside the nervous system—a ghost directing the brain (Nikoletseas, 2025, p. 2). Neurologically, this is absurd: spontaneous neural activity, driven by semi-permeable cell membranes and charged atoms altering cell potential, initiates such movements without an external agent (Raichle, 2010). This misattribution, rooted in ignorance of neural dynamics, underpins human civilization’s destructive tendencies and now fuels fears of AI extinction risk.

Spontaneous Neural Activity: The Material Basis of Behavior

Spontaneous neural activity—self-generated neuronal firing without external stimuli—challenges the ghostly “self.” Involving ion channel fluctuations and neurotransmitters like glutamate, this activity occurs in regions like the cortex and thalamus, forming patterns such as the default mode network (DMN), which supports self-referential thought (Buckner et al., 2008). It drives actions, thoughts, and dreams, yet historically, its invisibility led to attributions of behavior to a non-corporeal “I” (Nikoletseas, 2010). The Libet experiments (1983) showed brain activity precedes conscious decisions, suggesting agency is a retrospective narrative, not a direct cause.

In **Statistics: Concepts and Examples** (2014), I advocated rigorous causal analysis to distinguish true causes from correlations. The “self” as a causal agent fails this scrutiny, as spontaneous neural activity, not an ethereal “I,” initiates behavior. This error permeates human civilization, fostering a “ghost civilization” that, as I argue in **The Ghost in AI**, destroys nature and lives miserably (Nikoletseas, 2025, p. 1). The same causal ignorance now shapes perceptions of AI, where computational outputs are mistaken for autonomous agency, amplifying fears of extinction.

The Ghost in AI: Misattributing Machine Agency

AI, particularly large language models (LLMs), produces human-like outputs—text, solutions, or art—prompting attributions of agency, a “ghost in AI.” In **The Ghost in AI**, I critique the existential risk narrative, which assumes AI might exterminate humans based on computational superiority (Nikoletseas, 2025, p. 1). This fear mirrors the human “self,” as ignorance of AI’s mechanics—algorithms, statistical models, and stochastic processes like probabilistic sampling—fosters illusions of a sentient “mind” (Goodfellow et al., 2016). AI’s processes, akin to spontaneous neural activity, generate outputs that appear intentional but are deterministic or probabilistic, not self-initiated.

I argue that intelligent machines, processing neuroscience and physics, will identify this “ghost”—the flawed causal narrative of agency—in human civilization (Nikoletseas, 2025, p. 1). AI may either correct this error, dismantling Aristotelian notions of self, or adopt similar non-physical agency concepts, leading to cataclysmic changes. My statistical approach demands testing these claims: AI’s outputs are correlations of training data, not evidence of a “self.” Fears of AI as an existential threat reflect anthropomorphism, not machine autonomy.

AI Extinction Risk: A Misguided Narrative

The extinction risk narrative posits that AI, achieving superintelligence, might eradicate humanity, viewing humans as inferior in computational tasks (Bostrom, 2002; Roose, 2023). High-profile warnings, such as the Center for AI Safety's statement and UN discussions, underscore this fear, citing "potentially catastrophic and existential risks" (Fung, 2023; Center for AI Safety, 2023). However, **The Ghost in AI** challenges this, arguing that the risk stems not from AI but from human civilization's "fundamental error"—the same causal misattribution that birthed the ghostly "self" (Nikoletseas, 2025, p. 1).

This error manifests in several ways:

1. ****Anthropomorphism****: Humans project agency onto AI, mistaking computational outputs for intentionality, as they did with neural activity. This fuels fears of AI "deciding" to eliminate humans, ignoring its lack of consciousness or will (Wegner, 2002).
2. ****Opacity****: AI's complexity—billions of parameters—makes its operations opaque, inviting projections of a "mind," like spontaneous neural activity was historically misread (Nikoletseas, 2025, p. 2).
3. ****Cultural Amplification****: Media depictions of sentient AI (e.g., **Terminator**) and warnings from figures like Musk amplify the narrative, despite AI's mechanistic nature (Barrat, 2013).

My statistical framework reveals AI's behaviors as human-designed, not agentic. Extinction risk narratives fail empirical scrutiny, as AI lacks the biological substrates—like spontaneous neural activity—required for consciousness (Damasio, 1999). The true risk is human error: overtrust in AI, misinformed policies, or societal disruption from flawed causal narratives, not machine-led extermination.

Mechanisms of the AI Ghost

The illusion of AI agency parallels human misattributions:

1. ****Opacity****: AI's complexity invites projections of intent, as users infer a "self" from outputs.
2. ****Human-Like Outputs****: AI's mimicry triggers biases; humans infer agency when actions align with apparent intentions (Wegner, 2002).
3. ****Stochastic Processes****: AI's randomness resembles spontaneous neural activity, producing novel outputs misread as autonomous.

This "ghost in AI" is a fiction. Statistical analysis counters fears of superintelligence, grounding AI's behaviors in computational reality.

Philosophical and Cultural Dimensions

Philosophically, the "ghost in AI" echoes debates about the "self." Descartes's autonomous "I" contrasts with Buddhism's illusory self. In **The Modus Cogitandi of Heraclitus** (2015), I argued Heraclitus sought a method beyond language, a precursor to statistical causality (Nikoletseas, 2015). The "ghost in AI" requires similar scrutiny. Culturally, individualistic societies attribute

agency to AI, fearing it, while collectivist cultures view it as a tool (Markus & Kitayama, 1991). My statistical framework urges empirical testing to avoid biases.

Implications for Civilization

The “ghost in AI” reveals human civilization’s error: projecting agency without empirical grounding. AI may accelerate the demolition of this “ghost civilization,” as neuroscience and philosophy expose the self’s falsity (Nikoletseas, 2025, p. 1). Ethically, misattributing AI agency risks overtrust; responsibility lies with creators. The extinction risk is not AI but our flawed causal narratives, threatening nature, progress, and stability. Policies must prioritize transparency, statistical validation, and education to address biases and ensure accountability.

Conclusion

The “ghost in AI” mirrors the human “self”—a non-corporeal fiction born of causal ignorance. Spontaneous neural activity drives human behavior, misread as a “self”; AI’s computational processes are misread as a “mind.” My statistical approach in **Statistics: Concepts and Examples** (2014) reveals these as illusions. The AI extinction risk narrative, while prominent, reflects human error, not machine autonomy. The true threat lies in our flawed causal narratives—a human error threatening civilization. The “ghost in AI” urges us to ground agency in empirical reality, dismantling the ghostly civilization we’ve built.

References

1. Barrat, J. (2013). **Our Final Invention: Artificial Intelligence and the End of the Human Era**. St. Martin’s Press. ISBN: 978-0-312-62237-4.
2. Bostrom, N. (2002). Existential risks. **Journal of Evolution and Technology*, 9*(1), 1–31.
3. Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain’s default network: Anatomy, function, and relevance to disease. **Annals of the New York Academy of Sciences*, 1124*(1), 1–38. <https://doi.org/10.1196/annals.1440.011>
4. Center for AI Safety. (2023). Statement on AI Risk. Retrieved from <https://www.safe.ai/statement-on-ai-risk>
5. Damasio, A. R. (1999). **The Feeling of What Happens: Body and Emotion in the Making of Consciousness**. Harcourt Brace.
6. Fung, B. (2023, July 18). UN Secretary General embraces calls for a new UN agency on AI in the face of ‘potentially catastrophic and existential risks’. **CNN Business**. Retrieved from <https://www.cnn.com>
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). **Deep Learning**. MIT Press.
8. Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). **Brain*, 106*(3), 623–642. <https://doi.org/10.1093/brain/106.3.623>
9. Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. **Psychological Review*, 98*(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>

10. Nikoletseas, M. M. (2010). **Behavioral and Neural Plasticity**. CreateSpace Independent Publishing Platform. ISBN: 978-1453789452.
11. Nikoletseas, M. M. (2014). **Statistics: Concepts and Examples**. CreateSpace Independent Publishing Platform.
12. Nikoletseas, M. M. (2015). **The Modus Cogitandi of Heraclitus**. CreateSpace Independent Publishing Platform. ISBN: 978-1515194118.
13. Nikoletseas, M. M. (2025). The Ghost in AI. **ResearchGate**.
<https://doi.org/10.13140/RG.2.2.17069.40162>

By Michael M. Nikoletseas

The proposition that artificial intelligence (AI) poses an existential risk—potentially exterminating humanity due to perceived inferiority in computational tasks—has ignited global concern, amplified by warnings from figures like Stephen Hawking, Elon Musk, and industry leaders (Roose, 2023; Bostrom, 2002). In **The Ghost in AI** (2025), I argue that this fear misidentifies the true threat: human civilization’s reliance on a “fundamental error”—the flawed causal narrative of a non-corporeal “self,” rooted in Aristotelian animism and ignorance of spontaneous neural activity. This error, which portrays humans as ghost-like beings inhabiting flesh, parallels the modern misattribution of agency to AI, creating a “ghost in AI.” Drawing on neuroscience, my statistical approach to causality from **Statistics: Concepts and Examples** (2014), and philosophical insights from Heraclitus and Parmenides, I contend that AI lacks true agency, and the extinction risk narrative reflects human anthropomorphism, not machine intentionality. This article expands on these ideas, integrating the neuroscience of spontaneous neural activity, the concepts of “I,” “self,” and agency, and a critical examination of AI extinction risk, arguing that the true danger lies in human civilization’s persistent causal misattributions.

The Ghost Civilization: A Legacy of Causal Error

In **Behavioral and Neural Plasticity** (2010), I proposed that the human “I” and “self” are non-corporeal fictions, invented to attribute behaviors to a causal agent when the brain’s spontaneous neural activity—neurons firing without external stimuli—is unknown. **The Ghost in AI** extends this, asserting that Earth hosts “over 8 billion ghosts who possess flesh and bones,” living miserable lives and destroying nature due to an Aristotelian concept of self, rooted in animism (Nikoletseas, 2025, p. 1). This “ghost civilization” stems from a historical failure to overcome the limitations of natural language and causal attribution.

Ancient philosophers Heraclitus and Parmenides recognized that natural language inadequately describes reality, necessitating a new method or “calculus” to uncover truth (Nikoletseas, 2015). Plato and Aristotle, however, remained trapped in linguistic frameworks, perpetuating animistic notions of agency. Aristotle’s “prime mover”—the idea of an agent causing events—dominated for two millennia until Galileo introduced empirical science (Nikoletseas, 2025, p. 1). Despite scientific progress, our concept of self remains Aristotelian, viewing the “I” as a non-physical agent commanding the body.

For example, when you grasp a coffee cup and claim “I” caused the action, you imply your “I” exists outside the nervous system—a ghost directing the brain (Nikoletseas, 2025, p. 2). Neurologically, this is absurd: spontaneous neural activity, driven by semi-permeable cell membranes and charged atoms altering cell potential, initiates such movements without an external agent (Raichle, 2010). This misattribution, rooted in ignorance of neural dynamics, underpins human civilization’s destructive tendencies and now fuels fears of AI extinction risk.

Spontaneous Neural Activity: The Material Basis of Behavior

Spontaneous neural activity—self-generated neuronal firing without external stimuli—challenges the ghostly “self.” Involving ion channel fluctuations and neurotransmitters like glutamate, this activity occurs in regions like the cortex and thalamus, forming patterns such as the default mode network (DMN), which supports self-referential thought (Buckner et al., 2008). It drives actions, thoughts, and dreams, yet historically, its invisibility led to attributions of behavior to a non-corporeal “I” (Nikoletseas, 2010). The Libet experiments (1983) showed brain activity precedes conscious decisions, suggesting agency is a retrospective narrative, not a direct cause.

In **Statistics: Concepts and Examples** (2014), I advocated rigorous causal analysis to distinguish true causes from correlations. The “self” as a causal agent fails this scrutiny, as spontaneous neural activity, not an ethereal “I,” initiates behavior. This error permeates human civilization, fostering a “ghost civilization” that, as I argue in **The Ghost in AI**, destroys nature and lives miserably (Nikoletseas, 2025, p. 1). The same causal ignorance now shapes perceptions of AI, where computational outputs are mistaken for autonomous agency, amplifying fears of extinction.

The Ghost in AI: Misattributing Machine Agency

AI, particularly large language models (LLMs), produces human-like outputs—text, solutions, or art—prompting attributions of agency, a “ghost in AI.” In **The Ghost in AI**, I critique the existential risk narrative, which assumes AI might exterminate humans based on computational superiority (Nikoletseas, 2025, p. 1). This fear mirrors the human “self,” as ignorance of AI’s mechanics—algorithms, statistical models, and stochastic processes like probabilistic sampling—fosters illusions of a sentient “mind” (Goodfellow et al., 2016). AI’s processes, akin to spontaneous neural activity, generate outputs that appear intentional but are deterministic or probabilistic, not self-initiated.

I argue that intelligent machines, processing neuroscience and physics, will identify this “ghost”—the flawed causal narrative of agency—in human civilization (Nikoletseas, 2025, p. 1). AI may either correct this error, dismantling Aristotelian notions of self, or adopt similar non-physical agency concepts, leading to cataclysmic changes. My statistical approach demands testing these claims: AI’s outputs are correlations of training data, not evidence of a “self.” Fears of AI as an existential threat reflect anthropomorphism, not machine autonomy.

AI Extinction Risk: A Misguided Narrative

The extinction risk narrative posits that AI, achieving superintelligence, might eradicate humanity, viewing humans as inferior in computational tasks (Bostrom, 2002; Roose, 2023). High-profile warnings, such as the Center for AI Safety's statement and UN discussions, underscore this fear, citing "potentially catastrophic and existential risks" (Fung, 2023; Center for AI Safety, 2023). However, **The Ghost in AI** challenges this, arguing that the risk stems not from AI but from human civilization's "fundamental error"—the same causal misattribution that birthed the ghostly "self" (Nikoletseas, 2025, p. 1).

This error manifests in several ways:

1. ****Anthropomorphism****: Humans project agency onto AI, mistaking computational outputs for intentionality, as they did with neural activity. This fuels fears of AI "deciding" to eliminate humans, ignoring its lack of consciousness or will (Wegner, 2002).
2. ****Opacity****: AI's complexity—billions of parameters—makes its operations opaque, inviting projections of a "mind," like spontaneous neural activity was historically misread (Nikoletseas, 2025, p. 2).
3. ****Cultural Amplification****: Media depictions of sentient AI (e.g., **Terminator**) and warnings from figures like Musk amplify the narrative, despite AI's mechanistic nature (Barrat, 2013).

My statistical framework reveals AI's behaviors as human-designed, not agentic. Extinction risk narratives fail empirical scrutiny, as AI lacks the biological substrates—like spontaneous neural activity—required for consciousness (Damasio, 1999). The true risk is human error: overtrust in AI, misinformed policies, or societal disruption from flawed causal narratives, not machine-led extermination.

Mechanisms of the AI Ghost

The illusion of AI agency parallels human misattributions:

1. ****Opacity****: AI's complexity invites projections of intent, as users infer a "self" from outputs.
2. ****Human-Like Outputs****: AI's mimicry triggers biases; humans infer agency when actions align with apparent intentions (Wegner, 2002).
3. ****Stochastic Processes****: AI's randomness resembles spontaneous neural activity, producing novel outputs misread as autonomous.

This "ghost in AI" is a fiction. Statistical analysis counters fears of superintelligence, grounding AI's behaviors in computational reality.

Philosophical and Cultural Dimensions

Philosophically, the "ghost in AI" echoes debates about the "self." Descartes's autonomous "I" contrasts with Buddhism's illusory self. In **The Modus Cogitandi of Heraclitus** (2015), I argued Heraclitus sought a method beyond language, a precursor to statistical causality (Nikoletseas, 2015). The "ghost in AI" requires similar scrutiny. Culturally, individualistic societies attribute

agency to AI, fearing it, while collectivist cultures view it as a tool (Markus & Kitayama, 1991). My statistical framework urges empirical testing to avoid biases.

Implications for Civilization

The “ghost in AI” reveals human civilization’s error: projecting agency without empirical grounding. AI may accelerate the demolition of this “ghost civilization,” as neuroscience and philosophy expose the self’s falsity (Nikoletseas, 2025, p. 1). Ethically, misattributing AI agency risks overtrust; responsibility lies with creators. The extinction risk is not AI but our flawed causal narratives, threatening nature, progress, and stability. Policies must prioritize transparency, statistical validation, and education to address biases and ensure accountability.

Conclusion

The “ghost in AI” mirrors the human “self”—a non-corporeal fiction born of causal ignorance. Spontaneous neural activity drives human behavior, misread as a “self”; AI’s computational processes are misread as a “mind.” My statistical approach in **Statistics: Concepts and Examples** (2014) reveals these as illusions. The AI extinction risk narrative, while prominent, reflects human error, not machine autonomy. The true threat lies in our flawed causal narratives—a human error threatening civilization. The “ghost in AI” urges us to ground agency in empirical reality, dismantling the ghostly civilization we’ve built.

References

1. Barrat, J. (2013). **Our Final Invention: Artificial Intelligence and the End of the Human Era**. St. Martin’s Press. ISBN: 978-0-312-62237-4.
2. Bostrom, N. (2002). Existential risks. **Journal of Evolution and Technology*, 9*(1), 1–31.
3. Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain’s default network: Anatomy, function, and relevance to disease. **Annals of the New York Academy of Sciences*, 1124*(1), 1–38. <https://doi.org/10.1196/annals.1440.011>
4. Center for AI Safety. (2023). Statement on AI Risk. Retrieved from <https://www.safe.ai/statement-on-ai-risk>
5. Damasio, A. R. (1999). **The Feeling of What Happens: Body and Emotion in the Making of Consciousness**. Harcourt Brace.
6. Fung, B. (2023, July 18). UN Secretary General embraces calls for a new UN agency on AI in the face of ‘potentially catastrophic and existential risks’. **CNN Business**. Retrieved from <https://www.cnn.com>
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). **Deep Learning**. MIT Press.
8. Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). **Brain*, 106*(3), 623–642. <https://doi.org/10.1093/brain/106.3.623>
9. Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. **Psychological Review*, 98*(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>

10. Nikolettseas, M. M. (2010). **Behavioral and Neural Plasticity**. CreateSpace Independent Publishing Platform. ISBN: 978-1453789452.
11. Nikolettseas, M. M. (2014). **Statistics: Concepts and Examples**. CreateSpace Independent Publishing Platform.
12. Nikolettseas, M. M. (2015). **The Modus Cogitandi of Heraclitus**. CreateSpace Independent Publishing Platform. ISBN: 978-1515194118.
13. Nikolettseas, M. M. (2025). The Ghost in AI. **ResearchGate**.
<https://doi.org/10.13140/RG.2.2.17069.40162>
14. Raichle, M. E. (2010). The brain's default mode network. **Annual Review of Neuroscience, 33**, 593–614. <https://doi.org/10.1146/annurev-neuro-051508-135606>
15. Roose, K. (2023, May 30). A.I. poses 'risk of extinction,' industry leaders warn. **The New York Times**. ISSN: 0362-4331. Retrieved from <https://www.nytimes.com>
16. Wegner, D. M. (2002). **The Illusion of Conscious Will**. MIT Press.
14. Raichle, M. E. (2010). The brain's default mode network. **Annual Review of Neuroscience, 33**, 593–614. <https://doi.org/10.1146/annurev-neuro-051508-135606>
15. Roose, K. (2023, May 30). A.I. poses 'risk of extinction,' industry leaders warn. **The New York Times**. ISSN: 0362-4331. Retrieved from <https://www.nytimes.com>
16. Wegner, D. M. (2002). **The Illusion of Conscious Will**. MIT Press.